



*Congress of the United States
House of Representatives
Washington, D. C. 20515*

*Anna G. Eshoo
Eighteenth District
California*

September 20, 2022

Mr. Jake Sullivan
National Security Advisor
National Security Council
1600 Pennsylvania Avenue, NW
Washington, DC 20500

Dr. Alondra Nelson
Acting Director
Office of Science and Technology Policy
1650 Pennsylvania Avenue, NW
Washington, DC 20502

Dear Advisor Sullivan and Director Nelson,

I'm writing to express grave concerns about the recent unsafe release of the Stable Diffusion model by Stability AI. I strongly urge you to address this and similar unsafe releases using any authorities and methods within your power, including export controls, and request that you brief my office on any additional authorities the executive branch may need to address this issue.

On August 22, 2022, Stability AI released its open-source, text-to-image generation model called Stable Diffusion. Unlike OpenAI's DALL-E 2, Stable Diffusion's model is available for anyone to use without any hard restrictions. Predictably, Stable Diffusion was immediately misused after the model was released. Stability AI knew or should have known that Stable Diffusion would be misused and took no discernable steps to protect against these misuses before release. In one instance, Stability AI even provided further directions for how to misuse the model.

Following the open-source release of Stable Diffusion, photos of violently beaten Asian women generated by Stable Diffusion were posted in online chat rooms.¹ Reports also indicate that several 4chan threads have been dedicated to Stable Diffusion-generated pornography, some of which portray real people.² In a message posted to users of the Stable Diffusion Discord, Stability AI Founder and CEO Emad Mostaque said to Stable Diffusion users, "If you want to make NSFW [Not Suitable for Work] or offensive things make it on your own GPUs when the model is released."³ Mr. Mostaque then went on to tell users which GPUs were compatible with its model for the sake of using it to generate illicit content⁴, content Mr. Mostaque knew or should have known would likely include illegal content.

Unfortunately, the extent to which illegal or otherwise potentially dangerous images using Stable Diffusion were generated is unknowable due to its open-source nature,

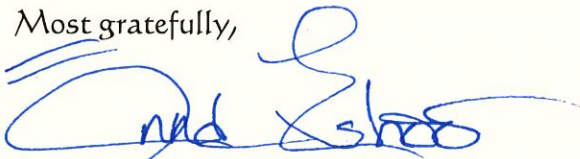
but it is both plausible and probable that pornographic images depicting real people under the age of 18 have been generated on individual users' computers and have created a market for Stable Diffusion-generated illegal depictions of minors, as well as other illegal content. While Stability AI's licensing terms do not provide for illegal content, the open-source release of the model provides for egregious dual-use applications. Stable Diffusion also includes a tool that attempts to detect and block offensive or undesirable images, but that tool can be easily circumvented using the open-source code. This means Stable Diffusion can be – and reportedly has been – used to create images that DALL-E 2 currently blocks⁵, including propaganda, violent imagery, pornography, images that potentially violate corporate copyright, and images used for disinformation and misinformation campaigns.⁶

Reporting suggests Stability AI released the unsafe model for funding purposes, as it is now in talks to raise capital and has cemented partnerships with “governments and leading institutions”⁷ and the model was reportedly released shortly after it was trained.⁸ I am an advocate for democratizing access to AI and believe we should not allow those who openly release unsafe models onto the internet to benefit from their carelessness. Democratizing access to AI may help alleviate incentives to release or deploy unsafe models⁹, and I've been leading this charge through my leadership in the AI Caucus, as well as through my legislation to develop a detailed roadmap for how the U.S. can build, deploy, govern, and sustain a national research cloud and associated research resources in order to make AI systems safer and more ethical by democratizing access to AI resources and testing.

While I commend Stability AI for its overall objective of democratizing access to AI, dual-use tools that can lead to real-world harms like the generation of child pornography, misinformation, and disinformation should be governed appropriately. In the same way that nuclear information and materials may lead to both the generation of energy and horrible atrocities, AI models similarly pose dual-use applications in a digital environment. We currently use export controls to control the release of various types of dual-use technical data, and I urge you to investigate the possibility of using such powers to control the release of unsafe dual-use AI models as well. In an increasingly digital world, we should increase our vigilance against digital harms to both individuals and society.

For all the reasons I've stated, I strongly urge you to address the release of unsafe AI models similar in kind to Stable Diffusion using any authorities and methods within your power, including export controls, and to brief my office on any additional authorities the executive branch may need to address this issue.

Most gratefully,



Anna G. Eshoo
Member of Congress

¹ Bazk T. Future, "Statement on Stable Diffusion," *Multimodal by Bazk T. Future*, August 27, 2022, <https://bakztfuture.substack.com/p/statement-on-stable-diffusion> [hereinafter *Future*].

² Kyle Wiggers, "Deepfakes For All: Uncensored AI Art Model Prompts Ethics Questions," *TechCrunch*, August 24, 2022, https://techcrunch.com/2022/08/24/deepfakes-for-all-uncensored-ai-art-model-prompts-ethics-questions/?utm_source=Center+for+Security+and+Emerging+Technology&utm_campaign=c4ad1997b6-EMAIL_CAMPAIGN_2022_09_08_12_51&utm_medium=email&utm_term=0_fcbacf8c3e-c4ad1997b6-438340193.

³ Future, *supra* note 1.

⁴ Ibid.

⁵ "Lessons Learned on Language Model Safety and Misuse" (OpenAI, March 3, 2022), <https://openai.com/blog/language-model-safety-and-misuse/>.

⁶ Benj Edwards, "With Stable Diffusion, You May Never Believe What You See Online Again," *Ars Technica*, September 6, 2022, https://arstechnica.com/information-technology/2022/09/with-stable-diffusion-you-may-never-believe-what-you-see-online-again/?utm_source=Center+for+Security+and+Emerging+Technology&utm_campaign=c4ad1997b6-EMAIL_CAMPAIGN_2022_09_08_12_51&utm_medium=email&utm_term=0_fcbacf8c3e-c4ad1997b6-438340193.

⁷ Kenrick Cai, "Startup Behind AI Image Generator Stable Diffusion Is In Talks To Raise At A Valuation Up To \$1 Billion," *Forbes*, September 7, 2022, <https://www.forbes.com/sites/kenrickcai/2022/09/07/stability-ai-funding-round-1-billion-valuation-stable-diffusion-text-to-image/?sh=41af5cdc24d6>.

⁸ Future, *supra* note 1.

⁹ Deep Ganguli et al., "Predictability and Surprise in Large Generative Models." In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022.